



DEFENCE



DÉFENSE

Managing Uncertainty in Digital Geographical Information: Final Report (U)

Sylvia Lam

Defence Research Establishment Ottawa

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-076
September 1999



National
Defence

Défense
nationale

Can:



Managing Uncertainty in Digital Geographical Information: Final Report (U)

Sylvia Lam
*EW Information Technology Group
Electronic Support Measures Section*

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-076
September 1999

Project
2JE14

Abstract

In recent years, data quality and uncertainty management has received strong interest from the geomatics community. The data quality issues related to the use of digital geographical information (DGI) are important because the migration from paper maps to digital databases has introduced a whole new paradigm on managing and using geospatial data. The Canadian Forces (CF) is a major producer and consumer of geographical data. Therefore, it is important for the CF to stay abreast of developments pertaining to data quality issues. This project is aimed at understanding the novel concepts in data quality and techniques in managing uncertainty in DGI. The project consists of many work units to explore the elements of data quality, methods of assessment, documentation, visualization, and uncertainty management techniques. The details of the experiments and results are described in five contractor reports. This document compliments the contractor reports by presenting the research materials at a conceptual level. The benefits of this project are a thorough understanding of the important issues related to data quality and firsthand experience with many of the data quality assessment and uncertainty management methods using military-supplied DGI. This provides the CF with expertise to respond to current and future requirements in this area.

Résumé

La qualité de l'information et la gestion de l'incertitude ont soulevé beaucoup d'intérêt de la part de la communauté des professionnels de la géomatique, au cours des dernières années. Les problèmes reliés à la qualité de l'information géographique numériques (IGN) sont importants puisque la migration des cartes géographiques en papier vers les bases de données numérique a introduit un tout nouveau concept en ce qui a trait à la gestion et l'usage de l'information géospatiale. La défense nationale (DN) est un producteur et consommateur majeur d'information géospatiale. Il est donc primordial pour la DN de se maintenir au courant des nouveaux développements dans la recherche sur la qualité de l'information. Ce projet a pour objectif d'amener la DN à comprendre les nouveaux concepts en qualité de l'information ainsi que les plus récentes techniques en gestion de l'incertitude en IGN. Le projet a exploré les éléments constituant la qualité de l'information, les méthodes d'évaluation de la qualité, de la documentation, de la visualisation ainsi que les techniques de gestion de l'incertitude. Les détails sur l'expérimentation et ses résultats sont présentés dans cinq rapports écrits par des contracteurs. Ce document complète les rapports des contracteurs en présentant le matériel de recherche à un niveau conceptuel. Les bénéfices retirés du projet consistent en une compréhension approfondie des problèmes reliés à la qualité de l'information, ainsi qu'une expérience de première main sur les méthodes d'évaluation de la qualité et de la gestion de l'incertitude utilisées avec des IGN militaires. Cela permet au DN d'avoir maintenant l'expertise nécessaire pour répondre aux besoins présents et futurs dans ce domaine.

Executive Summary

TR No. 1999-076

Sylvia Lam, Defence Research Establishment Ottawa

Managing Uncertainty in Digital Geographical Information: Final Report

Introduction

The migration from paper maps to digital spatial databases has introduced a whole new paradigm on using and managing geospatial data. This has prompted many research activities in identifying issues related to the proper use of digital geographical information (DGI) and in developing techniques to manage uncertainty related to DGI. The major concerns in DGI are data interoperability and data misuse.

The Canadian Forces (CF) is a major producer and consumer of geographical data. Therefore, it is important for the CF to stay abreast of developments pertaining to data quality issues. This project is aimed at understanding the novel concepts in data quality and techniques in managing uncertainty in DGI. It consists of many work units to explore the elements of data quality, methods of assessment, documentation, visualization and uncertainty management techniques, including the use of fuzzy sets. The details of the experiments and results are described in five contractor reports. This document compliments the contractor reports by presenting the research material at a conceptual level.

Results

This document provides a definition of uncertainty in DGI and examines the common approaches to manage uncertainty. These include uncertainty reduction and absorption, modeling and simulation. Summarized from literature research, this document provides a basic definition of the data quality elements that are required to fully describe the characteristics of a set of DGI. A good understanding of these basic elements allows researchers to develop assessment methods to measure them both quantitatively and qualitatively. This document also presents some of the better-defined methods and the results of applying these methods to a set of Digital Nautical Chart (DNC) data supplied by the CF Map Depot. Unfortunately, much of the data quality information pertaining to this data set is missing, and therefore reverse engineering is required to generate much of the needed parameters before one can apply the assessment methods. The lack of data quality information on DGI has been a hindrance to gaining strong attention from users and maintaining their enthusiasm.

Significance

The benefits of this project are a thorough understanding of the important issues related to data quality and first hand experience with many of the data quality assessment and uncertainty management methods using military-supplied DGI. This provides the CF with expertise to respond to current and future requirements in this area. This project was started as a technology-based project at Defence Research Establishment Valcartier. As the scientific authority has been subsequently deployed to Defence Research Establishment Ottawa, Section 5 of this document is devoted to relating data quality issues to electronic warfare applications.

Table of Contents

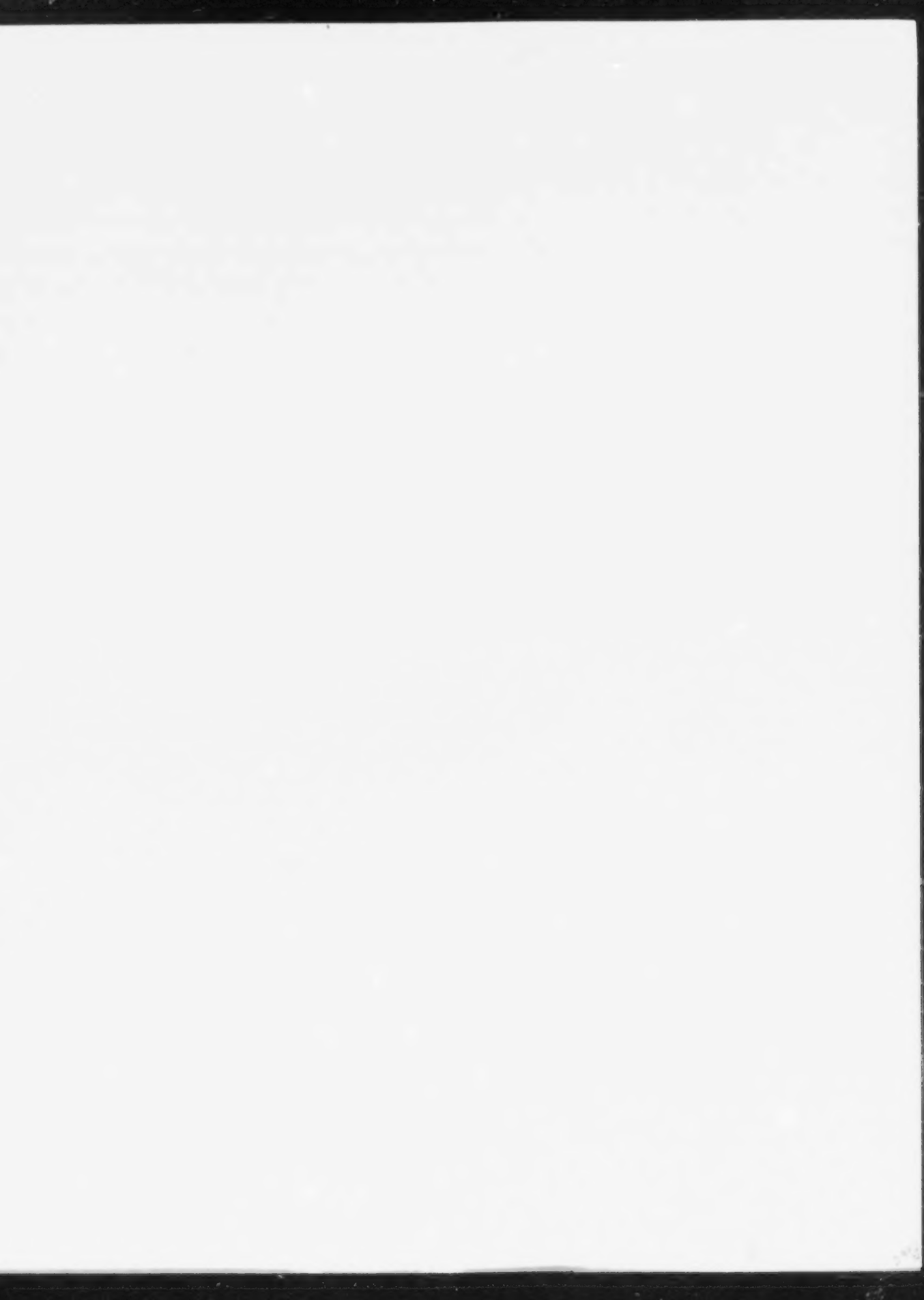
Abstract	iii
Executive Summary	v
Table of Contents	vii
List of Figures.....	ix
List of Tables	xi
1.0 Introduction.....	1
2.0 Elements of Data Quality in DGI	3
3.0 Data Quality issues in DGI	7
3.1 Novel Concepts in Data Quality	7
3.2 Assessing Data Quality Elements.....	9
3.2.1 <i>Data Quality Control for Paper Maps</i>	9
3.2.2 <i>Assessing the Quality of Digital Databases</i>	10
3.3 Documenting Data Quality Information	14
3.4 Visualization of Data Quality Information	16
3.5 Standardization in Geomatics.....	17
4.0 Managing Uncertainty in DGI	21
4.1 Uncertainty Reduction and Absorption	21
4.2 Uncertainty Management using Fuzzy Sets.....	23
4.2.1 <i>Data Queries</i>	24
4.2.2 <i>Data Modeling and Spatial Analysis</i>	25
5.0 Using Digital Geographical Information in EW	27
6.0 Conclusions	31

List of Figures

Figure 1. Disagreement between two digital databases on a stream intersection.....	3
Figure 2 A 1:1 million scale vector database overlaid on a 1:50,000 ADRG map image..	8
Figure 3 Best measurement between two line segments on a 1:1 million map.....	8
Figure 4 Hierarchy of different types of completeness.....	12
Figure 5 Some possible positions for an imprecise line.....	17
Figure 6 The set of <i>young</i> and <i>old</i> represented as (a) Boolean sets (b) fuzzy sets.....	23
Figure 7 (a) The fuzzy membership function for "about" 100 m.....	24
(b) A spatial search with a crisp search boundary.....	24
(c) A spatial search with a fuzzy search zone.....	24
Figure 8 The membership functions of (a) $A \cup B$ and (b) $A \cap B$	25

List of Tables

Table 1 Maximum computed standard deviation in planimetric diagnostic points for map ratings.....	10
Table 2 Example of a misclassification matrix.....	11
Table 3 Data quality evaluation of a DNC data set.....	13
Table 4 VMap Level 1 accuracy classification.....	29



1.0 Introduction

As geospatial data is incorporated into mainstream data processing, data quality and uncertainty issues in digital geospatial information (DGI) have gradually evolved from academic research topics to daily operational concerns. Many interpretations of the terms *data quality* and *uncertainty* exist in the open literature. In this report, we consider data quality an invariant characteristic of a data set, and refer to uncertainty as the condition when a decision-maker has doubts about the content of a data set. Therefore, the amount of uncertainty in a data set varies with decision-makers and the applications. For instance, a street map serves both a tourist and an emergency response team, yet it is unlikely that a tourist would worry too much about unlabelled one-way streets or misplaced fire hydrants.

Before the proliferation of desktop geographical information systems (GIS), geographical data was often collected, assembled and analyzed by professionals in mapping sciences. Nowadays, any user can easily manipulate DGI with a user-friendly desktop GIS. However, they may or may not understand the implications of the quality of the data set in use, or the consequences of transformation / conversion procedures. In some instances, convenience overrides data quality concerns when DGI are easily obtained on the Internet or bundled with GIS software. Blackmore [1] provided a classic example of misusing DGI: a digital data set of boundary coordinates for the British administrative district was created as a backdrop for thematic mapping purposes. The data set was intended for internal use and the applications of the data set do not require high accuracy. Nonetheless, due to its convenient software format, the data set became extremely popular. Soon secondary users found problems with point-in-polygon searches where some locations were found to be 2 or 3 km out into the North Sea.

The implications of data quality and uncertainty in DGI become more significant when used in mission-critical applications or when it involves the loss of lives and properties. Epstein et al [2] pointed out that the growing information industry is increasingly concerned with the exposure to litigation as their customers expect data from the computer to be "right", especially when these databases are sold at substantial prices.

The Canadian Forces (CF) is a major producer and consumer of DGI. The Mapping and Charting Establishment (MCE) supplies most of the DGI used by the CF. This includes raster and vector databases and some imagery. In cooperation with other NATO countries, MCE produces one of the world's most comprehensive vector-based digital geospatial data products, the Vector Smart Map (VMap). This product, along with other databases and tools, are being developed to support the initiative of the digitization of the battlefield. These databases serve a wide variety of users ranging from cockpit pilots to facility managers of military-owned properties. Being both a producer and a consumer of geospatial data, the CF must be aware of the concerns of data quality issues from both perspectives and participate in understanding and managing these issues.

J2 Geomatics Engineering takes a significant role in the development of standards in the military, national and international settings. The issues of data quality and uncertainty are subjects of discussion in all the standardization committees. The work presented in this report is performed under a technology-based project, work unit 2JE14, with participation from J2 Geomatics Engineering. The objectives of this project were to understand the fundamental concepts related to data quality, and to examine methods and techniques, especially fuzzy sets, to manage uncertainty arising from the use of DGI. Detailed descriptions on the experiments and results have been presented in five contractor reports [3 to 7]. This document complements the contractor reports by presenting the subject at a conceptual level, and refers to results and findings in the contractor reports for illustration and support.

The organization of this document is as follows: Section 2 introduces the basic elements of data quality in DGI. Section 3 examines the issues related to the documentation, assessment and visualization of data quality. Section 4 describes the approaches proposed by various researchers to manage uncertainty in digital geographical information. Section 5 examines data quality issues with respect to applications in electronic warfare. Section 6 presents the conclusions of the project.

2. 0 Elements of Data Quality in DGI

The migration from paper maps to digital databases has introduced new ways to manipulate and (re)use geographical data. The replacement of cartographers by empowered users has brought forth many concerns on the misuse or misinterpretation of spatial data and the result of data processing. Traditionally, a cartographer's responsibility includes collecting, editing and compiling data and producing maps. Maps were often end products of certain data compilations and map users had limited capability to manipulate information presented on paper. Nowadays, paper maps are being digitized and reused as data sources for other map compilations. Sometimes, even trained cartographers have difficulties understanding the implications of the "new-found freedom" provided by GIS. For instance, the proper way to integrate multiple-source, multi-scale DGI into a single database is yet to be established. Contractor Report Five [7] examined some of the problems associated with multiple-source data integration and proposed a method for such a process. Figure 1 shows a common problem in integrating different databases. Both Map 1 and Map 2 contain hydrographic features, i.e. streams and rivers, yet somewhere in the map there exists a 25-meter difference in a stream junction. To decide which map provides more accurate information requires knowledge of their respective data quality. This goes beyond simple spatial accuracy measurement because other data quality elements affect the overall accuracy of the map.

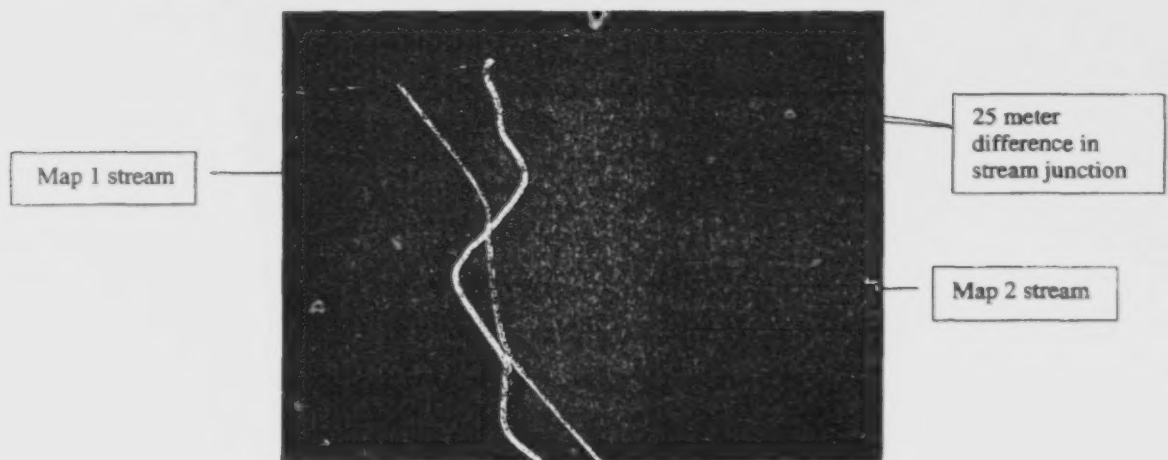


Figure 1. Disagreement between two digital databases on a stream intersection

Summarized from open literature, the information required to fully describe the quality of a digital geographical database are captured in seven elements:

1. Lineage
2. Spatial accuracy
3. Attribute accuracy
4. Temporal accuracy
5. Logical accuracy
6. Completeness
7. Semantic accuracy

A user survey [3] was carried out in this project to identify and prioritize users' concerns in the use of DGI. When asked to rank the importance of these seven elements, users consistently ranked spatial and attribute accuracies being the two most important on the list. Undoubtedly, these are two fundamental elements to the map users. Nevertheless, the response may also indicate the users' lack of familiarity with other elements. These seven elements are described in the following paragraphs.

Lineage. Lineage of a data set tells where the data comes from and the procedures that have been applied to the data to produce it. Information on lineage should include the source of the data, i.e. the source data set or signal from which the data set was formalized and the producer organization. The acquisition, compilation and derivation procedures should also be described. In addition, a data set often goes through conversions (e.g. projection, datum, format) to meet certain application requirements. These conversions could alter the content and structure of the data set and therefore, should be reported in the lineage documentation. Although lineage contains no direct measurement on any data quality parameters, it serves as a valuable source of information on data quality when other measurements are not available. Among all of the quality measures, lineage information is the easiest to collect and to carry over.

Spatial Accuracy. Spatial accuracy is a familiar term to most map users. Spatial accuracy tells the user how accurate is the location of a feature in absolute or relative terms. Spatial accuracy is usually described quantitatively with the level of confidence and the spatial reference system. For instance, Digital Terrain Elevation Data (DTED) Level 1 has an absolute horizontal accuracy of less than or equal to 50 meters 90% circular error with reference to the World Geodetic System (WGS 84). The absolute vertical accuracy of DTED is less than or equal to 30 meters 90% linear error with reference to Mean Sea Level (MSL). Spatial accuracy can be affected by the accuracy of measuring equipment, the resolution of the scanner used for digitization, the quality of the original maps used for digitizing and the skill of the human operators.

Attribute Accuracy. An attribute describes the characteristic of a certain location. It can be a characteristic of physical features on the ground (e.g. lakes) or a characteristic derived from abstract concepts (e.g. high-risk area). Both physical and abstract features can be subjected to error and uncertainty, with instrumentation errors and interpretation errors being the most common types. Some errors can be easily identified (e.g., forestland classified as suburban area); others are subject to interpretation. For instance, the natural

transition from one type of vegetation cover to another is often gradual and therefore difficult to draw as a crisp boundary on the map. As a result, the assessment of attribute accuracy may vary from one evaluator to another.

Completeness. Because maps are abstraction of the real world, only selected features can be presented in a certain compilation. The set of criteria used for feature selection and generalization is defined by an abstract model describing the universe of discourse. There are many interpretations of completeness. The most common interpretations are feature and attribute completeness. Feature completeness deals with the question of whether all features said to be there are present in the data set. Attribute completeness measures the number of missing values in the attribute tables of the database.

Many existing data sets do not have specific information on the model or criteria used for data collection and presentation. Often, this information could not be traced unless the databases were produced by major mapping agencies. This makes the assessment of completeness more challenging.

Logical Consistency. Logical consistency originated from the concept of database consistency. Like most textual databases, spatial databases must satisfy key constraints and integrity rules. In addition, spatial databases must satisfy topological constraints. Common examples of topological inconsistencies are undershoots, overshoots, missing reference points, missing nodes and pseudo nodes. Many professional GIS software packages have built-in functions to collect statistics and even rectify these inconsistencies. A logical consistency report provides the type of tests performed, the number of features checked and the specific problem related to each identified inconsistency. Most often, errors are easy to rectify and most data sets can be free of topological and database inconsistencies.

Temporal Consistency. Temporal information plays a strong role in the interpretation of data quality for a given data set. Though lineage provides some temporal information about the data set in terms of the production and handling of the database, it does not deal with more precise issues of the temporal aspects of the features. Most feature attributes change over time (e.g. deforestation, re-growth). Some features are susceptible to movement (e.g. landslide). Completeness takes on a different sense if time is taken into account. Logical consistency can be affected by updates. The concern in temporal aspects of geospatial information is its consistency. Inconsistency may arise when the data set consist of features collected or updated over different time frames and time scales. It can also be the result of multiple-source data integration. However, there is little consensus in how temporal information shall be represented. Therefore, the most common way is to describe temporal particulars of the features in a report.

Semantic Accuracy. Semantic accuracy refers to the meaning of the geographical objects rather than the geometrical representation. It is described as "...the quality with

which geographical objects are described in accordance with the selected model" [8]. This element of data quality is intended to help the user evaluate the fitness-of-use of a data set for a particular purpose. Many researchers consider semantic accuracy as a more detailed description of completeness, attribute accuracy and temporal accuracy, and therefore prefer to deal with semantics as part of other elements.

3.0 Data Quality issues in DGI

3.1 Novel Concepts in Data Quality

The introduction of digital geographical databases is accompanied by a set of new concepts in data quality measures. In the past, spatial accuracy was the only concern for most map users. In the digital paradigm, users need more information in order to fully understand the quality of the data set they are handling. A new set of data quality elements has been defined in Section 2. We examine other issues related to the use of DGI in this section.

DGI requires documentation on data quality and other information related to the use of the data set due to the wider sources of distribution. Traditional map users would assume the quality of the maps meet certain standards if they were produced by reputable institutions. Hence, little documentation about the maps was produced for the users. Today, DGI is available almost everywhere: on the Internet, bundled with software, through government agencies or private companies. Depending on the source of the data set, users may or may not get further support on the data sets to evaluate whether the data is appropriate for a particular application. For instance, DTED data supplied by the CF Map Depot is only accompanied by information on directory and file structure, the user's responsibility on the protection, control and distribution of the data and an address and phone number for general inquiries. No data quality information is provided, nor does the documentation explicitly refer to any other external documentation for more information on data quality (e.g. product specifications).

Only recently, meta data related to quality and lineage has been included with data sets being distributed by government agencies or reputable organizations. The United States is the only country to enforce the proper use of DGI by making it mandatory to have meta data accompany all geospatial data sets being transferred between government agencies.

Another new concept related to data quality deals with the presentation of DGI. Paper map users cannot easily change the scale of the maps but GIS users can zoom in and out at will. A 1:1 million scale map can be easily redrawn as 1:50,000 and overlaid with another 1:50,000 map. On the first map, a line 0.5 mm thick represents a width of 500 meters on the ground but the same line width on the second map represents 25 meters. Theoretically, a 0.5 mm thick line on a 1:1 million map should be drawn as a 10 mm thick on a 1:50,000 map. Yet, the users cannot tell one map from the other in a GIS because both maps were drawn with the same precision graphics (Figure 2). The problem worsens when users perform calculations or analyses on the features. Many GIS will provide a measurement with a precision to many digits after the decimal point. Such precision is misleading because the measurement between two linear features represented by 0.5 mm lines cannot be more accurate than 1 kilometer on a 1:1 million map. As illustrated in Figure 3, assuming the center line (dotted) is the position of a line displayed

on a map browser, the distance between two line segments measured on the display is d . On a 1:1 million map, the best measurement of the distance should be $d \pm 500$ meters.

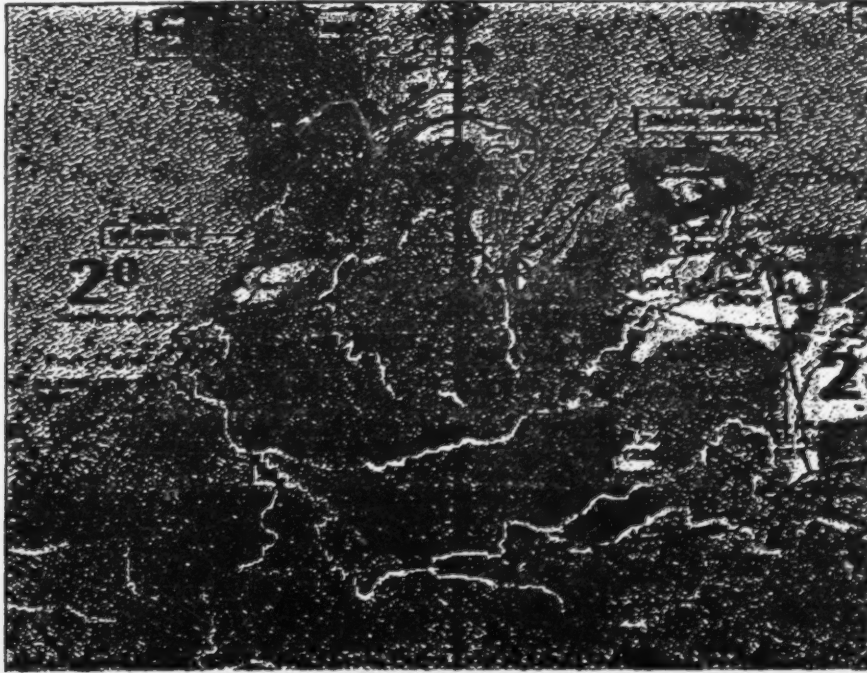


Figure 2. A 1:1 million scale vector database overlaid on a 1:50,000 ADRG map image. Blue and red lines represent streams and highways (1:1 million), respectively

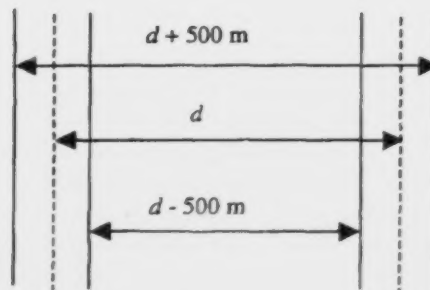


Figure 3. Best measurement between two line segments on a 1:1 million map is $d \pm 500$ meters

3.2 Assessing Data Quality Elements

Data quality in geospatial data is complicated by the nature of the data. Some features are inherently fuzzy and cannot be measured with high precision regardless of the quality of equipment, techniques and human interpreters. In addition, almost all geospatial databases are scaled abstraction of phenomena related to the Earth. Therefore, feature selection and generalization techniques are used to reduce the amount of data. Feature selection and generalization incorporate a certain amount of subjectivity from cartographers in order to retain useful information and to make maps legible. Furthermore, many feature parameters cannot be measured directly on the ground, therefore, remote sensing and photogrammetry techniques are required. Data quality can be affected by the type of techniques used and the skill of the human interpreter.

In general, accuracy of a data set is evaluated against another data set of higher accuracy or a *control data set*. Ideally, this control data set is the ground truth. Nonetheless, ground truthing is not always viable. Thus, the control data set could be another digital data set collected at a larger scale, collected by a more reputable organization, or collected with better equipment. Consequently, the data quality measures are relative to the control data set.

Most mapping agencies have their internal data quality control policy. Depending on the type of map being produced, their quality control procedures may emphasize different elements. For instance, spatial accuracy is of prime importance to topographic and hydrographic maps, while attribute accuracy and completeness are important to agricultural and forestry maps.

3.2.1 Data Quality Control for Paper Maps

Data quality control procedures on paper maps serve as a basis for developing procedures for digital databases. In this project, we have studied data quality control procedures of various map producers. Here we present an example from the production of the National Topographic Systems (NTS) paper maps [9].

The NTS establishes an acceptable quality level through stratified random sampling (i.e. by choosing map sheets based on the total number produced in any one batch) and statistical sampling of produced maps. The numbers of critical, major and minor defects in the map determine its quality rating (Table 1). Map sheets with more than 1 critical, 2 major or 20 minor defects cause them to be unacceptable.

At least 70 *diagnostic points* are sampled from each map sheet to determine its horizontal accuracy. At least 140 diagnostic points are sampled from each map sheet to determine its overall relief accuracy. Based on the standard deviation of the difference of these diagnostic points with the actual (ground truth) values for the position and height,

the map sheets are rated on a scale as shown in Table 1. A similar rating table exists for height accuracy with ratings ranging from 0 (best) to 4 (not determined). Complete specifications for Canadian NTS maps are given in [9], including a checklist for use by personnel carrying out the quality assurance testing.

Table 1. Maximum computed standard deviation in planimetric diagnostic points for map ratings [Source: ref. 9]

Rating	Scale of Map				
	1:25,000	1:50,000	1:100,000	1:200,000	1:250,000
A	12.5m	25m	50m	100m	125m
B	25m	50m	100m	200m	250m
C	50m	100m	200m	400m	500m
D	Poorer than C				
E	Not Determined				

3.2.2 Assessing the Quality of Digital Databases

Despite the differences in the components of error between paper map and digital databases, similar procedures are often used as a basis for evaluating the quality of DGI. Nonetheless, the difficulties with data quality elements in DGI are to define what to measure and how to measure it against a control data set. Since spatial accuracy evaluation procedures are well established in surveying engineering, we will focus our discussion on other elements.

Attribute Accuracy. Unlike spatial accuracy, many of the data quality elements do not have established references (e.g. WGS datums). Therefore, the assessment of these elements requires more conceptual definitions before one can establish an assessment method. For attribute accuracy, Goodchild [10] suggested that the scale of measurement is required to create a framework for the assessment of attribute accuracy. The scale of measurement is nominal, ordinal, interval and ratio, and most of the attributes use either the nominal or the interval scale. Attributes of the interval scale, e.g. mean temperature, can be assessed with statistical mean and standard deviation. Attributes of the nominal scale (e.g. land use classification) can be measured by comparing the classification found in a data set with the control data set. The differences in classification are usually communicated through a misclassification matrix (Table 2). The main diagonal contains the correctly classified cases, and numbers off the diagonal indicate the incorrectly classified cases. An overall measure of accuracy, the percentage correctly classified, can be calculated as the number of correct classifications divided by all cases. For the example given in Table 2, this percentage is $209/304 = 68.8\%$.

Table 2. Example of a misclassification matrix (Source: Ref. 10)

	A	B	C	D	E
A	80	4	0	15	7
B	2	17	0	9	2
C	12	5	9	4	8
D	7	8	0	65	0
E	3	2	1	6	38

Another index for measuring attribute accuracy is the Kappa index. This index takes into account that some correct classifications will occur by chance. It is calculated by:

$$\text{Kappa Index} = \frac{\sum_{i=1}^n C_{ii} - \sum_{i=1}^n \frac{C_{i.} C_{.i}}{C_{..}}}{C_{..} - \sum_{i=1}^n \frac{C_{i.} C_{.i}}{C_{..}}}$$

where n is the number of classes, i is the index over the class and the dots indicate summation over that index (e.g. $C_{.i} = C_{1i} + C_{2i} + C_{3i} + \dots + C_{ni}$)

There is also a distinction between the consumer's accuracy and the producer's accuracy. The consumer's accuracy is defined as the probability that a feature which appears to be class i is truly class i . The consumer's accuracy for class A, computed from Table 2, is $80/106$, or 75.5%.

$$\text{Consumer's Accuracy} = \frac{\sum_{i=1}^n C_{ii}}{\sum_{i=1}^n C_{i.}}$$

The producer's accuracy is defined as the probability that a feature that is truly class i will also appear as class i in the database. The producer's accuracy for class A, computed from Table 2, is $80/104$, or 76.9%.

$$\text{Producer's Accuracy} = \frac{\sum_{i=1}^n C_{ii}}{\sum_{i=1}^n C_{.i}}$$

Although the methods of assessment of attribute accuracy seem straightforward, the interpretation of the attribute values is sometimes subjective. As Goodchild [10]

describes "Two observers faced with identical information may even disagree as to whether a particular feature is one or the other." In addition, geographical features with indeterminate boundaries are difficult to separate using a crisp boundary. The question arises as to whether the error is in the position of the boundary or in the attribute labeling.

Completeness. Assessing data completeness requires a better understanding of what were the data collection rules, or the *abstract universe*, of a particular data set. Most references to completeness refer to feature or attribute completeness. Yet, Brassel et al [11] gives a more comprehensive hierarchy of different type of completeness (Figure 4).

Model completeness is used to evaluate the fitness for use. The abstract model of the data set is compared to that of the target application to verify if the required features and attributes are captured by the data collection rules. Data completeness consists of formal completeness and entity-object completeness. Formal completeness specifies if, or to what degree, the formal structure of a data set is complete. Formal completeness is required for the development of large databases or for the delivery of intermediate versions. Entity-object completeness uses the concepts in object-oriented technology to denote the class of geographical features (entity) and the occurrence of each instance of that feature (object). Thus, entity-object completeness measures to what degree objects belonging to a feature class are captured. Entity-object completeness is further extended into attribute completeness. Attribute completeness assesses how much information is captured to populate the database. For instance, an entity *bridge* is defined as any bridge that is longer than 10 meters and is built over a waterway. The attributes include its length, materials of construction, type of construction, and number of lanes. Entity-object completeness would measure how many bridges that fit the characteristics described above are in the abstract universe, and how many are actually captured in the database. Attribute completeness would measure how many of the attribute values are filled in.

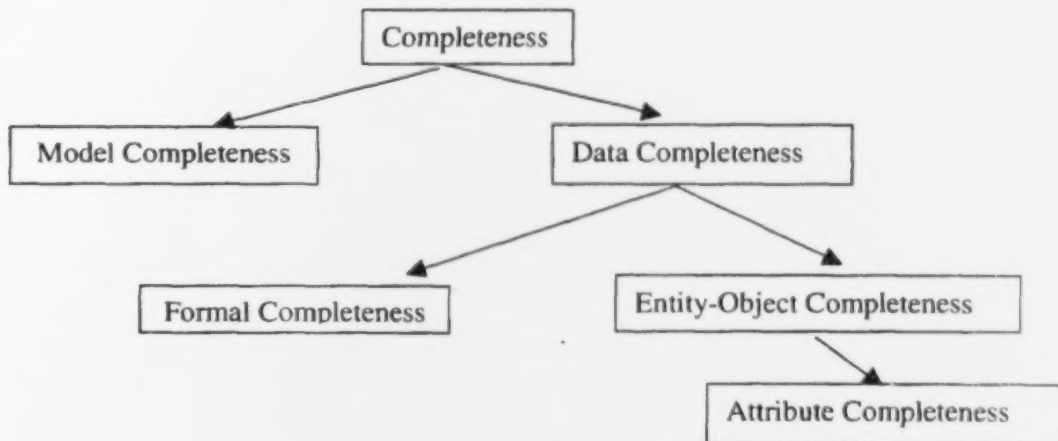


Figure 4. Hierarchy of different types of completeness (Source: ref. 11)

Logical Consistency. The concept of logical consistency originated with database technology to ensure database integrity. A consistent database must fulfill all consistency constraints to prevent invalid updates and to take proper actions whenever such an attempt is made. In addition to the logical consistency required for textual databases, spatial databases require two other types of logical consistency, i.e. topological consistency and scene consistency, to maintain their integrity. Kainz [12] provides a detailed discussion on the mathematical concepts behind logical consistency, namely Euclidean space using two and three-dimensional Cartesian coordinates, mathematical concept of topology and ordered sets. In short, topological consistency guarantees certain spatial properties in the databases. Some of these spatial properties are closed boundary of an area and planar topology. Scene consistency uses topology and ordered sets to enforce consistency within sets, or data elements. The consistency of a scene can be determined with the help of a connectivity matrix. An example of scene inconsistency would be a house built outside of its lot. Scene consistency is applied to avoid impossible or inconsistent operations.

Because logical consistency is governed by well-defined mathematical concepts and theories, many of the inconsistencies can be detected by software tools. Kainz [12] provided some examples of tests for logical consistency:

- Database consistency test
- Metric and incidence test
- Intersection test
- Euler equation test
- "Umbrella" test or node chaining
- Polygon chaining
- Order consistency test with depth-first search and transitive closure
- Topological scene consistency test

Temporal consistency. Every geographical feature has a temporal dimension associated with it. This temporal dimension is usually the time over which that feature is observed. Because data collection is usually carried out over a period of time, the temporal dimension may or may not be a concern, depending on the nature of the feature being observed. Temporal consistency becomes a concern when the database is partially updated over a long period of time and a temporally heterogeneous database results, or when the database is an integration of data collected from multiple sources.

Temporal consistency cannot be easily assessed unless a framework for modeling time is well established and well implemented in database management systems. Currently, the simplest way to record temporal information is to treat it as an attribute. Then, temporal accuracy can be assessed in the same way as attribute accuracy. Different definitions of time and types of temporal data structures have been proposed [13] but

have not been widely adopted. Guptill [14] emphasizes that temporal consistency should be considered together with other data quality elements. Often temporal inconsistency can account for positional inaccuracy, attribute error, data incompleteness and logical inconsistency.

In this project, we applied some of these assessment methods to evaluate a VMap 1:10,000 digital nautical chart (DNC) data set. Details of the process are described in Contractor Report Four [6]. The results are summarized in Table 3.

3.3 Documenting Data Quality Information

The need to create comprehensive data quality documentation for the general user has been in demand for years. Only in the past few years do we see some form of standards in reporting data quality information being put into practice. A well-known example is the Federal Geospatial Data Committee's (FGDC) Meta-data standard [15]. The FGDC meta-data standard aims at providing a common set of terminology and definitions for the documentation of DGI. The standard includes information needed to determine the availability of data, to evaluate fitness-for-use and to access and transfer identified data sets. The content of the FGDC standard amasses to a 90 page document. The basic content of the meta-data is expressed as follows:

Meta-data =

Identification_Information +
 0{Data_Quality_Information}1 +
 0{Spatial_Data_Organization_Information}1 +
 0{Spatial_Reference_Information}1 +
 0{Entity_and_Attribute_Information}1 +
 0{Distribution_Information}n +
 Metadata_Reference_Information

where $m\{term\}n$ denotes iteration, i.e. the term(s) enclosed is (are) repeated from m to n times. Each of the terms is expressed as group of data elements.

The *Data_Quality_Information* term in the above production rule can be further decomposed into :

Data_Quality_Information =
 0{Attribute_Accuracy}1 +
 Logical_Consistency_Report +
 Completeness_Report +
 0{Positional_Accuracy}1 +
 Lineage +
 (Cloud_Cover)

where (*term*) denotes an optional element.

Table 3. Data quality evaluation of a DNC data set (Source: Ref. 7)

DATA LAYERS										
COMPONENT	TEST INDICATOR	Cultural Landmarks	Earth Cover	Hydrography/Inland Water	Land cover	Limits	Aids to navigation	Obstructions/Port Facilities	Relief	
Lineage	1 Source date									
	2 other elements in FGDC metadata standard									
	3 A,B,C,D or E horizontal error N/C in	D			D			D	D	D
	4 0,1,2,3,4, vertical error N/C in			2						2
	5 1,2,3,4, horizontal error DIGEST	4			3	2		4	2	3
	6 1,2,3,4, Vertical error DIGEST			3						3
	7 Horizontal Circular error at 90%	113m			54 m	41 m	169 m	260 m	30 m	77 m
	8 vertical error at									6 m
	9 horizontal offset vector				53 m	19 m	88.4 m	85 m	72 m	55 m
	10 Vertical offset vector		0.1 m							0.2 m
Logical Consistency	11 EULER Count	1	1		1	1		1	1	1
	12 % non-node intersection	0%	0%		0%	0%		0%	0%	
	13 % inconsistent polygon chains	46%	0%		14%				79%	
	14 % inconsistent node chains	22%	0%		8%	2%		76%	15%	
attribute accuracy	15 % misclassified point features									
	16 % misclassified linear features: see table	table b	0%							
	17 % misclassified man-made linear feature	table b								
Completeness	18 % completeness of specific point/area feat	see table b								
	19 % completeness of specific of data	see table b	100%	100%	100%	100%	100%	100%	100%	100%
	20 % completeness of linear features	see table b						100%		
	21 percentage customer satisfaction		not measured							
semantic accuracy	22 standard deviation in days		not measured							
temporal accuracy										

Layer: Cultural Landmarks						
Component	Test INDICATOR	buildings	Roads	Trails	Ferry Crossings	Railroads/Tanks
attribute accuracy	15 % misclassified pt. Feature					42%
	16 % misclassified natural linear features	9%				
	17 % misclassified man-made linear features					
Completeness	18 % completeness of spec. pt/area features	11%		1%	50%	0%
	19 % completeness of data					62%
	20 % completeness of linear features			11%	6%	100%
						55%

The FGDC meta-data standard is often consulted by other standardization bodies including the committee of the Canadian Geospatial Data Infrastructure, the Digest Working Group and the International Standard Organization (ISO). In fact, ISO uses this standard as a base document for drafting the ISO Meta-data Standard 15046 Part 15, which is under revision for approval.

A challenge undertaken by the International Cartographic Association [16] is to develop an easy to understood index of spatial data quality. The index will be composed of a series of metrics that measure the elements of spatial data quality. The metrics will accompany digital data sets and will have meaningful recognition by the community of spatial data users. Undoubtedly, a thorough understanding of all aspects of the data quality elements is the basis for developing such an index. Nonetheless, the lack of standardized or generally accepted methods and metrics in measuring quality also make the task difficult.

One persisting obstacle in helping user to appreciate the documentation of data quality is that little data quality information exists in current databases and not enough information is being collected or recorded as the database grows. Even though most existing geospatial data interchange standards contain descriptions on how to store and present data quality information, these fields are often left blank. Reverse engineering these data quality parameters is a laborious task and the benefits may not always justify the cost. The users have no quality parameters to evaluate their own data sets or compute the index even when appropriate procedures exist. Consequently, the support from data users is less than enthusiastic.

3.4 Visualization of Data Quality Information

Besides textual reports, some research focuses on the graphical presentation of data quality information. One fundamental problem related to the visualization of data quality information is how to appropriately depict the information using various visual elements within the limited space available on the display surface [17]. A survey of visualization techniques is given in Hunter and Goodchild [18]. A natural choice is to use circles and ellipses to represent point uncertainty, and to use bands to represent line uncertainty. For area-based uncertainty, an uncertainty surface is commonly used. Variations of these techniques include the use of color, symbology, scale, texture and, more interestingly, animation.

Fisher [19] proposed the concept of data quality visualization through animation to show the uncertainty nature of dot maps, soil maps, and classified satellite imagery. A static display of point locations (e.g. location of radio emitters) gives an impression of precision even the locations of emitters are known to be estimates. To show the uncertain location of the locations, the points are randomly moved with their regions in real-time.

Soil maps are normally shown as choropleth maps with a uniform pattern applied to each zone. In reality, however, the zones could also belong partially to other classes since soil boundaries are not exact. To show this sampling uncertainty, the classification of the zones are randomly varied from one to another in real-time.

For satellite imagery, the likelihood of a pixel belonging to a certain class is computed. A similar real-time class switching technique can be applied, but the number of times a pixel stays in a class can be controlled by its likelihood of being in that class.

A dynamic technique for showing the uncertainty of lines can be developed in a similar way. Dutton [20] simulated line segments by connecting uncertainty endpoints. In the model, each point is assigned a circular locus of uncertainty that represents, for example, how far a point may wander from the center of a stream and still be in the water. There is a multitude of line segments that might be generated to connect adjacent points (Figure 5). On a dynamic display, the different connecting lines can be chosen at random, thus creating the perception of a fuzzy line.

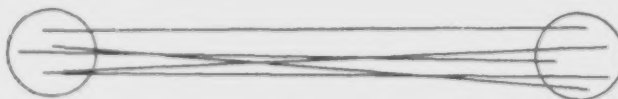


Figure 5. Some possible positions for an imprecise line

3.5 Standardization in Geomatics

One of the issues of data quality is data interoperability. To ensure DGI can be shared and reused effectively, a common approach to data quality management is required. Most standardization committees in geomatics address the issues of data quality. Among them are the NATO DIGEST Working Group, ISO TC/211 committee on geomatics, and the industry-led Open GIS Consortium. Defence Geomatics has been an active participant of these committees. These standardization bodies also work together to reduce duplication of work and to achieve harmonization among the standards.

DIGEST. The Digital Geographic information Exchange Standard is operated through the Digital Geographical Information Working Group (DGIWG) formed in 1990. DGIWG has eleven participating nations and two associated members. It started as a military standard and continues to support military requirements in its data products. The objectives of the DGIWG are "... to ensure that the geo community within defence can inter-operate, (and) to ensure that the geo community undertakes an element of burden

sharing." [21] More specifically, DGIWG is set out to identify requirements for DGI to support seven specific areas:

Terrain Analysis,
Transport and Logistics Planning,
Simulation,
Air Information,
Toponymy, and
exchange of Neutral Data.

DIGEST addresses both the exchange of geospatial data as well as the direct use of data products. It supports various types of data, levels of topology and encoding within a hierarchical standard structure. A current implementation of the DIGEST standard is the Vector Relational Format (VRF) which uses simple relational tables to organize data. A major effort is put on the production of standard-compliant databases: the VMap and DNC databases.

In the past few years, DGIWG has been focusing on the maintenance of the standard and in the production of large global databases (VMap levels 0 and 1). Current effort has also been put into the development of standards for raster data and imagery data.

ISO TC211. ISO TC211, formed in 1994, is an international effort to establish industrial and world standards in geomatics. It influences national standards of many countries and the GIS industry. The goal of ISO TC211 is to "...establish a structured set of standards concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth." [22] The standards specify the methods, tools and services required to acquire, manage, process, analyze, access, present and transfer geospatial data in digital format between different users, systems and sites. Existing information technology and other standards are linked to provide a framework for sector-specific applications using geospatial data. TC211 is potentially the most important component in the movement towards compatible geographical data standards because it provides a politically neutral forum for all proponents to align their respective standards.

The TC211 committee has set up working groups to address various topics. These topics include:

- issues related to framework and reference models
- issues related to geospatial data models and operators
- issues of geospatial data administration
- requirements for geospatial services
- profiles and functional standards

Five workgroups are formed to study and develop the standard. Each topic is further divided into many projects to focus on specific details. The TC211 committee meets twice a year to report on progress of projects and new resolutions. Many of the projects under the five topics are close to completion (i.e. near year 2000/2001). A notable part of their recent work is to invite co-operation with DGIWG, OpenGIS Consortium (below) and the international Hydrographic Bureau (IHB) to develop new work items as appropriate.

The OpenGIS Consortium. The OpenGIS Consortium, formed in 1994, is an open forum for the GIS community to contribute to the interoperability in geospatial data and operations. The objective of OGIS is "... to provide a framework for software developers to create software that enables their users to access and process geographic data from a variety of sources across a generic computing interface within an open information technology foundation." [22] The Open Geodata Interoperability Specification (OGIS) is developed under a non-profit, open consortium with participants from major sectors of the industry. OGIS plans to develop a detailed, platform- and product-neutral interface template for writing software that will inter-operate with other OGIS-compliant software written by different software developers.

The OGIS framework has three main parts: the Open Geodata Model (OGM), the OGIS Services Model (OSM) and the Information Communities Model (ICM). The OGM defines a basic set of geographic information types that can be used to model the geodata needs of more specific application domains. The OSM defines the set of services needed to access and process the geographic types defined in OGM. In addition, the OSM provides capabilities to share geospatial data within the communities of users that use different sets of geographic feature definitions. The ICM provides a schema for a community of people who already share a common set of semantics on geographic information. It also serves as a means for different communities to distribute and share data.

Although these three standardization bodies promote GIS interoperability, their objectives complement, rather than compete with, each other. DGIWG focuses on issues that are of particular interest to the NATO defence communities. ISO TC 211 provides an international infrastructure framework for world-wide interoperability. OGC focuses on the development of specifications on software technology that will have strong influence on future GIS systems. Members in these three bodies maintain close communications with each other to avoid duplication of effort and ensure harmonization on standardization policies.

4.0 Managing Uncertainty in DGI

Uncertainty describes the condition when a decision-maker has doubt about the nature or extent of what is known at the time he makes the decision [2]. Uncertainty in decision-making can be classified into three types [24]: 1) uncertainty due to imperfect knowledge of the environment, 2) future uncertainty due to the consequences of choices made in the present, and 3) uncertainty as to the appropriate value to be placed on personal judgement.

The work presented in this document focuses on issues related to the first type of uncertainty, i.e. uncertainty due to imperfect knowledge in geospatial databases and the reality it represents. The previous two sections have examined the new issues related to data quality in DGI. These issues directly affect the amount of uncertainty a decision-maker has to face when using DGI. Well-documented DGI will assist the decision-maker tremendously in assessing and reducing the amount of uncertainty he has to deal with.

Bédard [25] pointed out that a common approach to managing uncertainty is to first reduce the uncertainty to an acceptable level, then to absorb the rest of it. Uncertainty absorption refers to the situation where the decision-maker or an organization is ready to accept the consequence caused by the uncertainty (e.g. through financial compensation). Uncertainty reduction can be achieved by acquiring more information from other sources, and through modeling and simulation to evaluate the associated risk. Remaining uncertainty is usually absorbed by the decision-maker or his organization, or through third-party liability insurance. Some map producers are prepared to compensate their users for harm resulting from the use of erroneous information. The various ways to document, visualize and assess data quality help to reduce uncertainty and to evaluate the magnitude of the risk. Qualitative descriptions of data quality provide the decision-maker with a better understanding of the data set, while quantitative measurements (e.g. spatial accuracy) given in data quality reports can be used as parameters for modeling and simulation.

4.1 Uncertainty Reduction and Absorption

When perfect information is not available or not economically feasible to obtain, other methods can be used to estimate the amount of uncertainty in question. In aerotriangulation [26], for example, uncertainty is usually estimated by gathering errors in films, cameras, stereo-plotting instruments, instrument operators, earth curvature and map projections. However, the interaction of the different error sources could be complex and the modeling of the combined result is non-trivial. Another example of uncertainty modeling is given in Thapa and Bossler [27]. They have investigated the error sources of mapping and have illustrated the technique of combining errors for map production.

Interpolation and simulation techniques are other means to estimate the amount of uncertainty in geospatial data. Interpolation methods such as Kriging [28] are convenient

ways of estimating uncertainty based on known measurements at given locations. Image classification techniques also generate estimate of uncertainty as a by-product of processing the data [29]. Simulation is another technique that can be combined with other methods to estimate uncertainty. Zhou [30] proposed a combination of simulation and interpolation methods to improve the uncertainty estimation on elevation data due to attribute (elevation) and planimetric locations.

Several assumptions on spatial data sampling are often used in generating error models, and Goodchild [10] cautioned readers on their validity. Firstly, one often assumes data sampling errors are randomly distributed over the area of interest. Secondly, one assumes that sample points are representative of all the points on the map. However, in many situations, convenient test points are chosen to reduce the cost and to facilitate access or to obtain a more precise measurement. Thirdly, a particular type of error model (e.g. a normal distribution) is assumed to be a good predictor of the probability of the type of errors being assessed but not all errors share the same type of probability distribution. Finally, spatial correlation and class dependency, i.e. how the error in one class affects the accuracy of another class, are often overlooked.

When uncertainty in DGI cannot be completely eliminated, information providers must be willing to provide some kind of assurance to encourage the use of information they produce. Some information providers, in many cases government agencies, issue guarantee statements about the contents of information being correct and will compensate users for any harm resulting from the use of erroneous information. Often, these data sets are recognized as "official" or "certified" version which implies the acceptance by the producer as being reliable and that other competing data sets do not carry the same level of integrity [2]. In the Global Geospatial Information Infrastructure initiative [31], a similar approach is being used to provide assurance to the users: the CF evaluate and test software tools and data sets for their appropriateness for military applications, and distribute certified versions of these software and data sets to users. This certification procedure also ensures that the same analysis (e.g. line-of-sight computation) used in different applications produces consistent results.

Liability insurance on information databases can be considered a novel approach to uncertainty absorption. Liability insurance is a growing business as information is being traded as a commodity. Uncertainty about the quality and appropriateness of information is what information providers and users seek to guard against [2]. Epstein also provided a more detailed discussion on the mechanisms for insuring geographical information. In summary, the insurer will require the following information about the data producer to evaluate the risk of underwriting:

- background information on the information producer,
- the producer's quality control and assurance measures and its compliance to recognized standard such as ISO 9000,

- the producer's compliance to established industrial standards on formatting/transfer,
- meta data on quality of information,
- a certified copy from the supplier and securely held in escrow by an independent third party.

It is also expected that to gain insurance protection, producers perform work strictly according to the legal, professional and industrial standards, and provide sufficient documentation to demonstrate that their products have been produced to the best of their ability and knowledge.

4.2 Uncertainty Management using Fuzzy Sets

The essence of fuzzy sets [32] is the ability to express partial membership of a set. While Boolean logic allows only single membership in a class, fuzzy sets allow partial membership. For instance, in Boolean logic, a 40 year-old person can only belong to either set *old* or the set *young*. In fuzzy sets this person can belong to both sets with respective membership values, say 0.5 in the set *young*, and 0.5 in the set *old* (Figure 6). Fuzzy sets thus provide more flexibility in expressing concepts that are difficult to classify using Boolean sets.

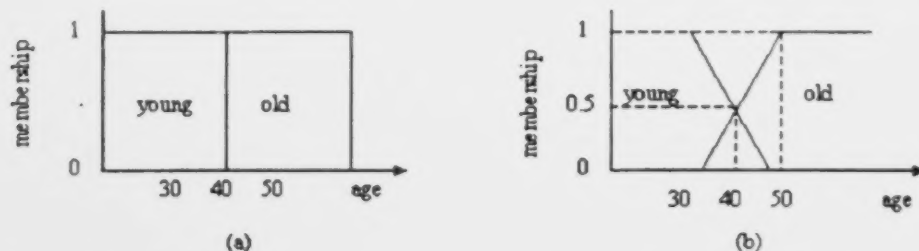


Figure 6. The set of *young* and *old* represented as (a) Boolean sets (b) fuzzy sets

In the past years, fuzzy set theory has found its way to many GIS applications. Proponents of fuzzy sets found the concept very appropriate to handle uncertainty in geospatial data because of the inherent fuzziness of many geographical features (e.g. ecological zones), or the fuzziness due to data collection criteria, scaling, generalization and other cartographic techniques. Fuzzy set-based techniques have been applied to geospatial query, data modeling and processing. One popular use of fuzzy set-based techniques is in satellite image analysis. Fuzzy set classifiers can be developed to provide optimal classification of features. Examples of other applications include environmental assessment [33] and wildlife management [34].

4.2.1 Data Queries

The implementation of fuzzy queries does not require data sets to be fuzzy. Therefore, it is one of the easiest ways to incorporate fuzzy set-based techniques into a GIS system. Examples of which can be found in [29, 34]. Many spatial queries involve searching for features with some criteria. An Electronic Warfare (EW) analyst may have obtained information from an intercept conversation that Unit "X" is *about* 100 meters from a major road intersection. In a conventional query, the analyst would issue a query to search for a radio emitter within 100 meters of the intersection, which automatically eliminates any emitters outside the 100-meter radius. This could risk eliminating a possible match that is slightly outside the search area (Figure 7b). In a fuzzy query, the analyst can define a fuzzy search zone of *about* 100 meters (Figure 7a). The system will then search for emitters that are within the fuzzy zone, and provide a membership value for each emitter indicating how well each emitter matches the search criteria (Figure 7c).

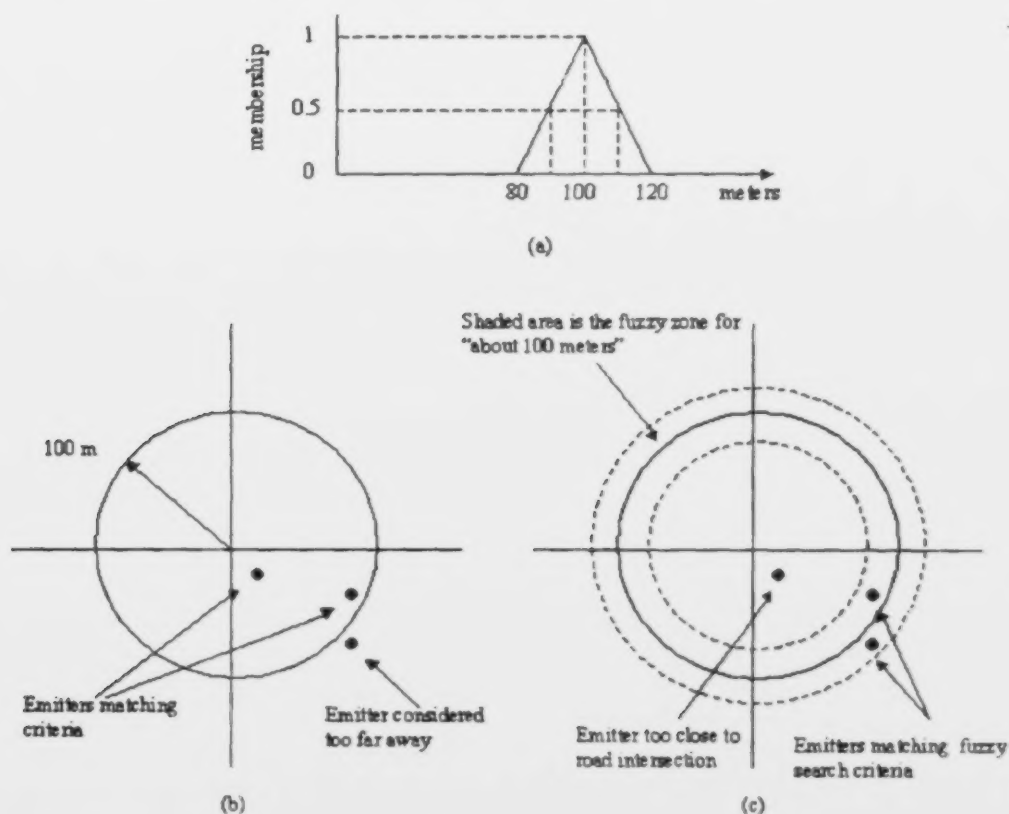


Figure 7. (a) The fuzzy membership function for "about 100 m". (b) A spatial search with a crisp search boundary (c) A spatial search with a fuzzy search zone

4.2.2 Data Modeling and Spatial Analysis

Fuzzy sets have been widely applied to the modeling of geographical regions with indeterminate boundaries. Leung [35] is one of the first publications to provide a comprehensive study on fuzzy sets applied to goespatial analysis. He defined a fuzzy region as a core and a boundary, with fuzziness often occurring at or around the cores and boundaries. Morris and Petry [36] applied these concepts to modeling the air pollution zone for a city. They defined any location within 30 units of the city centre as definitely within the air pollution zone. Locations greater than 30 units and less than 60 units have membership values which linearly relate to distance from the city centre. This simple example illustrates the concepts of core and boundary and the fuzzy zone as discussed in Leung. The membership function is given as follows:

$$\begin{aligned} f_{A^*}(V) &= 1 && : |V - (450, 500)| < 30; \\ f_{A^*}(V) &= \frac{60 - |V - (450, 500)|}{30} && : 30 < |V - (450, 500)| < 60; \\ f_{A^*}(V) &= 0 && : 60 \leq |V - (450, 500)|. \end{aligned}$$

where V is a map space $= [0, 1000]^2$ and $A^* \subset V$ is a fuzzy feature representing an air pollution danger zone around a city at location $(450, 500)$.

A set of basic operations has been defined for fuzzy sets by Zadeh [32]. They are the *union* and *intersection* operations. Figure 8 shows graphically the membership functions for $A \cup B$ and $A \cap B$. The functions are defined as follows:

Union:

$$\mu_{A \cup B}(x) = \max [\mu_A(x), \mu_B(x)], \forall x \in X$$

Intersection:

$$\mu_{A \cap B}(x) = \min [\mu_A(x), \mu_B(x)], \forall x \in X$$

where μ denotes the membership value.

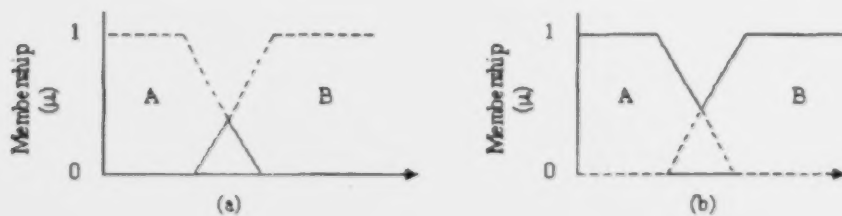


Figure 8. The membership functions (dotted lines) of (a) $A \cup B$ and (b) $A \cap B$

Many researchers have extended this basic set of operations by relaxing certain conditions or imposing new conditions to handle more complex manipulations. Operators such as *complement*, *difference*, *addition*, *subtraction*, *multiplication* and *division* by a real number are discussed in Leung. Compared to classical set theory, fuzzy set theory is still at its developmental stage and subject to further discussion and research. Nonetheless, fuzzy sets provide us with an interesting alternative to model and process imprecise data.

5.0 Using Digital Geographical Information in Electronic Warfare

DGI has gradually found its way into various military operations. The most popular DGI used in EW operations are DTED and ADRG maps. DTED provides elevation data to perform terrain analysis such as radio propagation modeling, while ADRG provides thematic coverage (i.e. roads, rivers, towns, etc.) for situation analysis. Presently, most situation analysis work is performed manually and ADRG maps are sufficient for this purpose. Nonetheless, the increasing demand for automation in EW analysis will eventually require more sophisticated DGI to support computer-aided analysis. As most EW activities happen at the operational level, many analyses require medium (1:250,000) to large scale data (1:50,000 and larger). The implications of data quality become more significant as the scale of maps gets larger because more details are expected at a higher precision in the database.

EW situation analysis requires various types of geographical information. The most common data sets being used are ADRG, DTED, and more recently, VMap datasets. VMap provides information in vector-relational format. Main features include hydrographic, railroad and highway networks, land cover, political boundaries and populated areas. When used together with combat and intelligence information collected on the battlefield, VMap provides necessary geographical information to support situation analysis and to compute mobility and radio propagation. Since these three data sets are prime sources of geographic information for EW analysis, users should be familiar with the quality of these datasets.

ADRG/CADRG. Firstly, users of ADRG should be aware that ADRG is being replaced by Compressed ADRG (CADRG) and new systems should consider the CADRG format for digital image of paper maps. According to MIL-C-89038 [37], the major differences between the two are the sampling rate (from 100 microns to 150 microns in CADRG), color coding (from 16.7 million to a maximum of 216 colors) and frame divisions (to enable direct use for avionics computers). The source graphics for CADRG are either produced by NIMA or other sources listed in [37]. Processing applied to the source graphics include spatial reduction with filtering, vector quantization for image compression, and color quantization and reduction. Spatial reduction also employs edge sharpening and smoothing algorithms to remove downsampling artifacts.

MIL-C-89038 does not describe specifically the errors associated with the processes applied to the source data. Nonetheless, the specification noted that radiometric colors on the images are representations of the colors in the source map products. The transformation process in CADRG emphasizes image contrast, clarity and legibility over exact color fidelity, and most of the color lost should be insignificant for military operations. It also indicated that the difference in quality between ADRG and CADRG is insignificant despite the lower sampling rate on CADRG's images. The accuracy of both ADRG and CADRG is dependent upon the accuracy of the source data, i.e. the original paper charts and maps. Information on the source data could be found in the source

support data accompanied with each CADRG volume. The source support data contains various information pertaining to the use of the data. This includes information on accuracy, original projection and datums of the source graphic. The vertical datum remains the same as the source graphic. The horizontal datum is converted to WGS 84 datum. Users will notice that latitude/longitude readings on the images do not correspond to the gridlines on the image graphics if the original datum was not WGS 84. Projection distortion in the East-West direction is present due to a stretch at the poleward latitude and a compression at equatorward latitude.

Since ADRG/CADRG are images of paper maps, emphasis is placed on spatial accuracy as demonstrated by the description provided in the documentation. Should the users wish to know the data collection criteria on the original paper maps, it is the user's responsibility to trace the source of information. In addition, many of the data quality elements discussed in Section 2 are not completely relevant in ADRG/CADRG. Since there exist no formal databases in ADRG, data completeness and logical consistency become irrelevant. Temporal inconsistency would be mainly due to inconsistent updating of multiple map sheets rather than information within each map sheet.

DTED. Since DTED provides only terrain elevation information, the focus of quality description is on its horizontal and vertical accuracy. According to MIL-PRF-89020 [38], the absolute horizontal accuracy is less than or equal to 50 meters, 90% circular Error (CE) WGS. The absolute vertical accuracy is less than or equal to 30 meters 90% linear error (LE) to Mean Sea Level (MSL). The relative horizontal and vertical accuracy over a one degree cell is 30 meters 90% CE WGS and 20 meter LE MSL, respectively. Each elevation point represents the true value referenced to MSL datum with a precision to the nearest meter. Each DTED level 1 file is organized in a one degree by one degree cell. An accuracy section is provided within each cell file to allow further accuracy information for a sub-region to be specified, though this section is not filled in for unclassified distributions.

According to the MIL-PRF-89020, information on database maintenance date, product specification, the digitizing collection system, compilation date and the like are provided on magnetic tape format but not on CD-ROM.

As with ADRG, logical consistency and completeness are irrelevant in DTED. DTED does allow missing values in the data file and are coded as -32767. Temporal consistency cannot be easily assessed and the existing file structure does not provide such information.

VMap. Although VMap is not yet widely used in military operations due to its limited availability, it should be considered a promising source of geographical information in the future, especially in support of situation analysis. Unfortunately, at present, the availability of VMap at medium and large scale is rare. Current VMap databases exist in 1:1 million scale (world-wide coverage) and for selected areas in 1:250,000 and larger.

The plan for VMap production is to have world-wide coverage at 1:250,000 and selected areas at larger scales. Nonetheless, operational requirements will dictate the availability.

According to the specification MIL_V-89033 [39], accuracies of VMap Level 1 (1:250,000) are classified as follows:

Table 4. VMap Level 1 accuracy classification

Absolute Horizontal Accuracy (class)	90% Circular Error	Absolute Vertical Accuracy (class)	90% Linear Error (contour interval)
1	125 m	1	0.5
2	250 m	2	1.0
3	500 m	3	2.0
4	> 500 m	4	> 2.0

Each VMap volume has several documentation files describing different information related to data quality. The files include lineage information, a data quality table, and data quality coverage. The lineage information and the data quality table describe the whole distribution volume. Information such as sources of data, reference datums, spatial accuracy, attribute and data completeness and logical consistency can be found in the data quality table. The data quality coverage provides another means to describe data quality of a coverage (i.e. layer of information) down to individual point, node, line or area feature. Despite these means for reporting data quality, they are not fully utilized at present. Many of the fields in the quality tables are either empty or contain generic information. The data quality coverage is rarely used. It is expected that these means for reporting data quality will eventually be utilized when the focus of the DIGEST committee switches from their present objective of data production to quality improvement.

To ensure easy access and exchange of data, EW software systems that require DGI should be designed to comply with military and/or international standards, and system designers should be aware of current developments in uncertainty management techniques which could improve EW analysis tools.

6.0 Conclusions

Novel concepts and fundamental issues related to data quality and uncertainty management in DGI have been presented. The basic elements for describing data quality in DGI have been described, and the issues related to the documentation, assessment and visualization of data quality elements have been discussed. The work of DGWIG, ISO/TC211, and OGC have also been introduced. These standardization committees will have significant influence on the future of geographical data exchange and software development in both civilian and military communities.

In Section 4, the concept of uncertainty was defined and related to the use of DGI. Different approaches have been proposed by various researchers to manage uncertainty in DGI. These include uncertainty reduction and absorption, modeling and simulation and the use of fuzzy set-based techniques. Section 5 presented the data quality issues with respect to the support of electronic warfare operations. The requirements for automated analysis tools for EW analysts have created the need to supplement the ADRG map images with more intelligent geographical database such as the VMap products. The sources of data quality information for ADRG, DTED and VMap were presented. They will provide important information for the design and performance evaluation of software tools for EW operations.

Current and future work in this area is diverse. Presently, effort in the geomatics community, many members of which are data producers, has been focusing on the definition, documentation and presentation of data quality information to support data and software interoperability. Much of the work is led by standardization committees. The user community tends to focus research work on the development of uncertainty management tools and modeling techniques to deal with the issues. Since DND is both a producer and a consumer of DGI, participation in research and development on both sides are needed. J2 Geomatics continues to be a strong participant and leader in the development of standardization at various levels of government to develop an infrastructure to achieve software and data interoperability in the geomatics community. Because of the wide range of applications, individual users should study the impacts of data quality and uncertainty pertinent to their applications.

References

- [1] Blackmore M. (1985). High or Low Resolution? Conflicts of Accuracy, Cost, Quality and Application in Computer Mapping. *Computer & Geoscience* Vol. 11, 345-348.
- [2] Epstein, E.F., Hunter, G.J. & Agumya, A. (1998). Liability Insurance and the Use of Geographical Information, *Int. J. Geographical Information Science*, Vol. 12, #3, 203-214.
- [3] WaterMark Industries Inc. (1995a). Managing Uncertainty in Digital Geographical Information: Survey of Technology, Phase I report. (DREV Contract W7701-4-1824/01-XSK). Defence Research Establishment Valcartier, Quebec, Canada.
- [4] WaterMark Industries Inc. (1995b). Managing Uncertainty in Digital Geographical Information: Survey of Technology, Phase 2 report, (DREV Contract W7701-4-1824/01-XSK). Defence Research Establishment Valcartier, Quebec, Canada.
- [5] WaterMark Industries Inc. (1996). Managing Uncertainty in Digital Geographical Information: Proof of Concept, Phase I report, (DREV Contract W7701-5-1725/01-XSK). Defence Research Establishment Valcartier, Quebec, Canada.
- [6] WaterMark Industries Inc. (1997). Managing Uncertainty in Digital Geographical Information: Proof of Concept, contractor report, (DREV Contract W7701-5-1725/01-XSK). Defence Research Establishment Valcartier, Quebec, Canada.
- [7] WaterMark Industries Inc. (1998). Data Quality Visualization, Final Report, (DREV Contract W7701-7-0884/001/XSK). Defence Research Establishment Valcartier, Quebec, Canada.
- [8] Salge, F. (1995). Semantic Accuracy in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York.
- [9] Energy, Mines and Resources Canada (1994). Accuracy. Canada Center for Mapping, Surveys Mapping and Remote Sensing Sector, DRAFT standard.
- [10] Goodchild M. (1995). Attribute Accuracy in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York.
- [11] Brassel K.B., Bucher F., Stephas E.M. and Vekovski A. (1995). Completeness in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York.
- [12] Kainz W. (1995). Logical Consistency in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York.

- [13] Guptill, S.C. (1995). Temporal Information in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York.
- [14] Langran, G. (1993). Time in Geographic Information systems. Taylor & Francis, Washington, DC.
- [15] FGDC, FGDC-STD-001-1998 (1998). Content Standard for Digital Geographic Meta-data. Federal Geospatial Data Committee, Washington D.C.
- [16] Veregin H. & Hargitai P. (1995). An evaluation matrix for geographical data quality in *Elements of Spatial Data Quality*, Eds. S.C. Guptill and JL Morrison, Elsevier Science Inc., New York, USA.
- [17] Tutfe, E.R. (1983). The Visual Display of Qualitative Information. Graphics Press, Cheshire, Connecticut.
- [18] Hunter, G.J. & Goodchild, M.F. (1993). Managing Uncertainty in Spatial Databases: Putting Theory into Practice, *URISA Proceedings*, Vol. 1, 1-14.
- [19] Fisher, P. (1992). Real-time Randomization for the Visualization of Uncertain Spatial Information, *Proceedings of the 5th Int. Symposium on Spatial Data Handling*, Charleston, South Carolina, Vol. 2, 491-494.
- [20] Dutton G. (1989). Modeling Locational Uncertainty in Hierarchical Tessellation in *Accuracy of Spatial Databases*, Eds Goodchild, MF and Gopal, S. Taylor and Francis, London, 125-140.
- [21] DIGEST Information page (Online), NIMA.
<http://www.tnmpo.nima.mil/guides/df/digest.html> (August 5, 1999).
- [22] ISO/TC 211, Home Page <http://www.statkart.no/isotc211/> (August 5, 1999).
- [23] Open GIS Consortium Inc, The OGIS Guide (Online). Maryland MA. Home Page <http://www.opengis.org/> (August 5, 1999).
- [24] Zwart, P. (1994). Land Information Systems and the Policy Process. PhD Thesis. University of Tasmania, Hobart, Australia.
- [25] Bedard, Y. (1987). Uncertainties in Land Information Systems Databases. *Proceedings of Auto-carto 8*, 175-184, Baltimore, Maryland.
- [26] Slama, C.C. (1980). Manual of Photogrammetry, Forth Edition. American Society of Photogrammetry.

- [27] Thapa K. and Bossler, J. (1992). Accuracy of Spatial Data used in Geographic Information Systems, *Photogrammetric Engineering & Remote Sensing*, Vol. 58, #6, 835-841.
- [28] Goodchild, M.F. (1992). Geographical Data Modeling, *Computer & Geoscience*, Vol. 18, #4, 401-408.
- [29] Goodchild, M.F., Montello, D.R., Fohl, P. and Gottsegen, J. (1998). Fuzzy Spatial Queries in Digital Spatial Data Libraries, *Proceedings, IEEE World Congress on Computational Intelligence*, 205-210.
- [30] Zhou, F.Q. (1995). Uncertainty in Geographical Information Systems, PhD thesis, Dept. of Geodesy and Geomatics Engineering, University of New Brunswick, New Brunswick, Canada.
- [31] Mckellar, D.G. (1995). The Global Geospatial Information and Services Initiative. J2 Geomatics Engineering. <http://132.156.33.161/GGIS/GGISdoc.html>, (Dec. 1995).
- [32] Zadeh, L.A. (1965). Fuzzy Sets, *Information and Control*, Vol. 8, 338-353.
- [33] Lein, J.K. (1992). Modeling Environmental Impact using an Expert-Geographic Information System. *GIS/LIS Proceedings*. San Jose, California, Vol. 1, 436-444.
- [34] Bock, W., and Salski, A. (1997). A Fuzzy Knowledge-based Model of Population Dynamics of the Yellow-necked Mouse (*Apodemus flavicollis*) in a Beech Forest. *Ecological Modeling*, Vol. 108, 155-161.
- [35] Leung, Y. (1988). Spatial Analysis and Planning under Imprecision, Elsevier Science Publishers B.V, The Netherlands.
- [36] Morris A. and Petry F.E. (1998). Design of Fuzzy Querying in Object-Oriented Spatial Data and Geographic Information Systems, *Proceedings, Conference of the North American Fuzzy Information Processing Society, NAFIPS, IEEE*, 165-169.
- [37] NIMA (1994). MIL-C-89038, Military Specification, *Compressed Arc Digitized Raster Graphics*, Mail Stop A-13, 8613 Lee Highway, Fairfax, VA 22031-2137.
- [38] NIMA (1996). MIL-PRF-89020, *Performance Specification, Digital Terrain Elevation Data*. Mail Stop A-13, 8613 Lee Highway, Fairfax, VA 22031-2137.
- [39] NIMA (1995). MIL-V-89033, Military Specification, *Vector Smart Map (VMap) Level 1* Mail Stop A-13, 8613 Lee Highway, Fairfax, VA 22031-2137.

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA (Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)		
1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence Research Establishment Ottawa Department of National Defence Ottawa, Ontario Canada K1A 0Z4	2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable) UNCLASSIFIED	
3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.) MANAGING UNCERTAINTY IN DIGITAL GEOGRAPHICAL INFORMATION: FINAL REPORT (U)		
4. AUTHORS (Last name, first name, middle initial) LAM, SYLVIA		
5. DATE OF PUBLICATION (month and year of publication of document) SEPTEMBER 1999	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.) 35	6b. NO. OF REFS (total cited in document) 39
7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.) DREO REPORT		
8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include full address.) Defence Research Establishment Ottawa Department of National Defence Ottawa, Ontario Canada K1A 0Z4		
9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant) 2JE14	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)	
10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.) DREO TR 1999-076	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)	
11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification) ☒ (X) Unlimited distribution ☐ () Distribution limited to defence departments and defence contractors; further distribution only as approved ☐ () Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved ☐ () Distribution limited to government departments and agencies; further distribution only as approved ☐ () Distribution limited to defence departments; further distribution only as approved ☐ () Other (please specify):		
12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). however, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)		

13. **ABSTRACT** (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

(U) In recent years, data quality and uncertainty management has received strong interest from the geomatics community. The data quality issues related to the use of digital geographical information (DGI) are important because the migration from paper maps to digital databases has introduced a whole new paradigm on managing and using geospatial data. The Canadian Forces is a major producer and consumer of geographical data. Therefore, it is important for the CF to stay abreast of developments pertaining to data quality issues. This project is aimed at understanding the novel concepts in data quality and techniques in managing uncertainty in DGI. The project consists of many work units to explore the elements of data quality, methods of assessment, documentation, visualization, and uncertainty management techniques. The details of the experiments and results are described in five contractor reports. This document compliments the contractor reports by presenting the research materials at a conceptual level. The benefits of this project are a thorough understanding of the important issues related to data quality and firsthand experience with many of the data quality assessment and uncertainty management methods using military-supplied DGI. This provides the CF with expertise to respond to current and future requirements in this area.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

DIGITAL GEOGRAPHICAL INFORMATION
DATA QUALITY
UNCERTAINTY

The Defence Research
and Development Branch
provides Science and
Technology leadership
in the advancement and
maintenance of Canada's
defence capabilities.

Leader en sciences et
technologie de la défense,
la Direction de la recherche
et du développement pour
la défense contribue
à maintenir et à
accroître les compétences
du Canada dans
ce domaine.

#512333



www.crad.dnd.ca

